

REC'D 11 APR 2003

WIPO PCT

Ref.6 PCT/PTO 15 JUL 2004

PCT/JP 03/00261

NOT AVAILABLE COPY

本 国 特 許 庁  
JAPAN PATENT OFFICE

15.01.03

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2002年 1月15日

出願番号

Application Number:

特願2002-042355

[ST.10/C]:

[JP2002-042355]

出願人

Applicant(s):

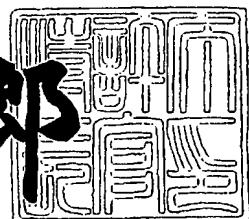
鈴木 俊明

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

2003年 3月25日

特許庁長官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3010562

【書類名】 特許願  
【整理番号】 TSJT02001  
【提出日】 平成14年 1月15日  
【あて先】 特許庁長官  
【国際特許分類】  
【発明者】  
【住所又は居所】 神奈川県藤沢市川名849の10クリオ藤沢参番館50  
6  
【氏名】 鈴木 俊明  
【発明者】  
【住所又は居所】 神奈川県鎌倉市玉繩5-12-9  
【氏名】 田中 順治  
【特許出願人】  
【住所又は居所】 神奈川県藤沢市川名849の10クリオ藤沢参番館50  
6  
【氏名又は名称】 鈴木 俊明  
【電話番号】 0466-28-8179  
【提出物件の目録】  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1

【書類名】 明細書

【発明の名称】 S N P (S i n g l e   N u c l e o t i d e   P o l y m o r p h i s m : 一塩基多型または一塩基多型となる座位) のタイピング・データおよびハプロタイプ解析を用いた、疾患易罹患性または薬剤応答性の特定方法

【特許請求の範囲】

【請求項 1】

疾患易罹患性または薬剤応答性に関連する S N P の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で S N P のタイピングを行う事によって、前記関連する S N P を特定することを特徴とする S N P 特定方法。

【請求項 2】

マーカーとなる S N P を推定する事で疾患易罹患性または薬剤応答性に関連する S N P の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で S N P のタイピングを行う事によって、前記関連する S N P を特定することを特徴とする請求項 1 記載の S N P 特定方法。

【請求項 3】

ハプロタイプ解析を用いてマーカーとなる S N P を推定する事で疾患易罹患性または薬剤応答性に関連する S N P の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で S N P のタイピングを行う事によって、前記関連する S N P を特定することを特徴とする請求項 1 記載の S N P 特定方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

S N P 機能解析プロセスの効率化と精度の向上、および製薬企業などでの開発新薬の治験における S N P 機能解析手法を応用した薬剤応答性の特定方法

【0002】

【従来の技術】

従来の疾患易罹患性や薬剤応答性に関連する SNP を特定する SNP 機能解析プロセスでは、コスト上の理由によって、解析したい SNP を数十から数千程度に絞り込んでから wet プロセス（注1）である SNP のタイピングを行っている。（図1中の「従来の SNP 機能解析のプロセス・フロー」に示す。）

これは、ヒトが持つ約300万の全 SNP について TaqMan 法（注2）でタイピングをした場合のコストが約2億円（サンプル一人当たりの試薬コスト）で、SNP 機能の統計的な解析に必要な数百から数千のサンプルのについてこの全 SNP のタイピングを行えば数百億円から数千億円の非現実的なコストと大規模な解析施設等のリソースが必要となるからである。

この為、通常の SNP 機能解析プロセスでは、タイピングする SNP（以下、「タイピング SNP」と称す）を限定し、1千から1万程度の SNP に絞り込んで機能の解析を行う。

しかしながら、疾患易罹患性や薬剤応答性と SNP の関連の有無は、その SNP をタイピングした結果から統計的に解明する以外に方法はない。この為、最終的に関連が解明される「目標」 SNP（注3）は、予め、タイピング SNP として1千から1万程度の SNP のグループに含まれて選定されていなければならない。これらの SNP が選定からもれた場合には、解析で関連の有る SNP の見つからず、解析プロセスを再度タイピング SNP グループの選定からやり直さなければならない。

タイピング SNP を選び出す従来のやり方は、研究者が論文等の文献やゲノム関連のデータベース等を検索し、機能が既に解明しているヒト以外のゲノムと類似性したヒトの遺伝子の機能を予測するホモロジー検索等の手法を用いている。

しかしながら、これらのゲノム情報には、ヒト・ゲノムの機能が完全に記載されていない。この為、この SNP 機能解析プロセスの効率を決定するタイピング SNP を選び出すステップ、つまり如何に高い確率で「目標」 SNP を予測出来るか否かは、研究者個人の経験とスキル、そして偶然の要素に大きく依存している。

SNP 機能解析プロセスの問題のもう一つにデータの品質が有る。

SNP 機能解析では、或る形質の発現の有無（例、作用や罹患の有無）によって

区分されるサンプル・グループの間でSNPのタイピングを行い、両グループにおける各SNPのアレルの頻度を統計的に解析し、その形質を発現させる要因となるSNPを特定する。つまり、wetプロセスのタイピング・データに品質に問題があった場合、そのデータを基にしたSNP機能解析解析の結果は不正確なものになる。

この問題は、SNPのタイピングは人間が介在するプロセスである事に起因する。SNP機能解析プロセス固有の品質の問題であるコンタミネーション、サンプルや試薬の取り違え等のオペレーション上のケアレス・ミス、これらデータ品質低下の原因となる要素の多くが人的なものであり、これもまた研究者個人の経験とスキルに大きく依存している。

(注1) wetプロセスとは、SNPのタイピングを行うプロセス。現在のTaqMan法では、プレート上で血液等の遺伝子サンプルと試薬を反応・ハイブリダイゼーションさせ、その結果を光学的に測定し、最終的にサンプル個々がそのSNPで取るアレルのタイピングを行う…以上の工程をここではwetプロセスと称する。特定されたタイピングのデータの統計解析は、wetプロセスには含まれない。

(注2) 蛍光標識したアレル特異オリゴとTaq DNAポリメラーゼによるPCR反応とを利用したタイピング方法

(注3) 「目標」SNPまたは「目標」となるSNPとは、疾患易罹患性や（開発新薬の）薬剤応答性の要因となるSNP、および疾患易罹患性や薬剤応答性の指標となるSNP、以上2つの何れかに該当するものを意味する。SNP機能解析の目的は、これらのSNPを特定する事である。

### 【0003】

#### 【発明が解決しようとする課題】

しかしながら、以上の従来技術によれば、タイピングの前に「目標」SNPを予測して、これらを含んだ数百から数千程度のSNPのグループを適切かつ確実に選択する事は困難であり、また、SNPタイピング・データの品質を低下させるwetプロセス中の人的エラーであるサンプルや試薬の取り違えやコンタミネーションの発生を防止する事も極めて困難である。

そこで、この発明は、マーカーとなる SNP の推定とその近傍 SNP の詳細なタイピングを繰り返す事で、「目標」 SNP が存在すると思われる塩基配列領域を段階的に絞込み、最終的に「目標」 SNP を効率良く特定する方法を提供することを課題とする。

加えて、wet プロセスにおけるデータ品質の要因となるオペレーション上のケアレス・ミスを防止するプロセス管理手法、コンタミネーション等によって汚染されたデータを統計解析の前に排除する機能を提供することを課題とする。

#### 【0004】

##### 【課題を解決するための手段】

以上の課題を解決するために、請求項 1 の発明は、疾患易罹患性または薬剤応答性に関連する SNP の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で SNP のタイピングを行う事によって、前記関連する SNP を特定することを特徴とする SNP 特定方法である。

また、請求項 2 の発明は、マーカーとなる SNP を推定する事で疾患易罹患性または薬剤応答性に関連する SNP の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で SNP のタイピングを行う事によって、前記関連する SNP を特定することを特徴とする SNP 特定方法である。

また、請求項 3 の発明は、ハプロタイプ解析を用いてマーカーとなる SNP を推定する事で疾患易罹患性または薬剤応答性に関連する SNP の解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込み、品質管理された工程で SNP のタイピングを行う事によって、前記関連する SNP を特定することを特徴とする SNP 特定方法である。

#### 【0005】

##### 【発明の実施の形態】

この発明の一実施形態である、疾患易罹患性または薬剤応答性に関連する SNP 特定のフローを、図 1 に示す。

このプロセス・モデルでは、以下 8 つの工程を実施する事によって、SNP のタ

イピングを行う「走査領域」から段階的に絞り込んで、最終的に開発新薬の薬剤応答性の有無と関連する「目標」SNPを特定する。

#### 前段階1：決定の対象となる開発新薬の決定

製薬企業等によって開発された新薬の内、開発新薬の薬剤応答性（=作用または副作用の有無）を SNP によって検定したものを選択する。又、このプロセス・モデルでは、これ以外に、例えば疾患易罹患性等を対象として、SNPとの関連を調べる事も可能である。

#### 前段階2：解析するサンプルの収集

SNP機能解析では、或る形質の発現の有無によって区分されるサンプル・グループの間で、SNPデータを比較し、その形質を発現させる要因となる SNP を特定する。例えば、糖尿病等について疾患易罹患性と SNPとのを調べる場合には、糖尿病患者のグループと、controlされたグループの間で、各 SNP のアレル頻度を統計的に分類する。この時、controlグループとして、糖尿病に罹患していないグループ、もしくは（糖尿病罹患の有無にかかわらず）無作為に抽出された平均的なサンプル・グループ、この2つの何れかを用いる。

開発新薬の薬剤応答性に関連する SNP を特定する場合、作用もしくは副作用のあったグループと、これが無かったグループについて、段階1以降の解析を行う。無作為に抽出された平均的なサンプル・グループが存在する場合、「もしくはこれらに該当する外部の SNP データが利用できる場合には、前述の2グループと合わせた3つのグループの間で、データの比較と解析を行うと、更に有効な解析を行う事が可能となる。

#### 段階1：「走査領域」の決定

このプロセス・モデルでは、この段階1から5までを1つのサイクルとしてこれを繰り返す事によって、初期の大まかな「走査領域」（注4）からより局所的な「走査領域」へと段階的に絞り込みを行う。最後のサイクルでは、「走査領域」中の存在する全 SNP のタイピング・データを解析する事で、最終的に「目標」SNPを特定する。

図2に示されるように、最初のサイクルにおける「走査領域」は、染色体等の大まかなレベルで規定する。これは、現状でも染色体等レベルの大まかな機能が解

明されているからである。また、これより詳細な、例えば遺伝子レベルで初期の「走査領域」を設定しても可である。

2回目以降のサイクルでは、（段階5から再度この段階1へ戻るので）段階5で連鎖不平衡が確認された領域を新しい「走査領域」として設定する。この時、絞り込まれた新しい「走査領域」は、前回の「走査領域」の数分の1から數十分の1となる。（次のサイクルで、具体的にどの程度絞り込むかについては、下記で述べるSNP間の連鎖不平衡の強度等によって異なる。）

（注4）「走査領域」とは、「目標」SNPの存在を調べる=走査する領域であり、ヒト・ゲノムの塩基配列上の連續した領域である。この領域の物理的な長さは、段階的に絞り込まれるために可変である。

#### 段階2：タイピングSNPの決定

前の段階で設定された「走査領域」中から、「タイピング」するSNPのグループを選定する。

これらグループに含まれるSNPは（遺伝子部位の機能やエクソン・イントロン等の区分に留意せず）任意に選ぶ事が出来るが、隣接するSNP座位の間隔が出来る限り等間隔となるように選択を行う。これは、この一連の解析でSNP座位間の連鎖不平衡を間接的に観測する為、連鎖不平衡に大きな影響を及ぼすSNP座位間の物理的距離の差による誤差を排除する為である。（図2を参照）

解析できるSNP間の連鎖不平衡は、SNP座位間の物理的距離が1～10万塩基程度と考えられる。この為、最初の「走査領域」が含まれるSNPが10万程度の染色体の全長である場合ならば、初回のタイピングは1000SNP程度を行う事が望ましい。

2回目以降のサイクルでは「走査領域」は段階的に絞り込まれ、「走査領域」の物理的距離は短くなり、その範囲に含まれるSNP数は減少する。タイピングSNPは、この「走査領域」の中から、数十から最大でも数百までの範囲で選択する。

#### 段階3：wetプロセスによるSNPタイピング

選択されたSNPグループについて、各サンプルのSNPタイピングをTaqMPCR法等によって行う。このタイピング工程におけるコンタミネーション

やサンプルの取り扱いによるデータの誤差を防ぐ為に、1) バーコードを用いたサンプル・試薬チューブおよびアッセイ用プレート等の世代管理、2) 「ハーディ、ワインバーグ平衡」を用いたタイピング・データの検定、以上を行う事でタイピング・データの品質保証を行う。

#### 1) バーコードを用いたサンプル・試薬チューブおよびアッセイ用プレート等の世代管理

タイピングの工程における最も基本的で多い間違いは、サンプルや試薬を間違えて取り扱う事である。現行のS N P 解析のプロセスでは、最終的にアッセイを行うタイピング機器に使用するアッセイ・プレートを生成するまでにも、幾つかの中間プレートを生成する。その為、これらのプレートについても正しいプレートから生成されているかの

サンプル：下部に二次元バーコードを刻印されたサンプル・チューブ、およびこのサンプル・チューブが96本まで格納出来るサンプル・ラックによって管理される。

サンプル・ラック内の各サンプル・チューブの配置は、サンプル・ラックの底面からラック内の各サンプル・チューブのバーコードをスキャナーで読み取り、この配置をデータとして管理する。又、サンプル・ラック自身にもバーコードが添付される。

試薬：下部に二次元バーコードを刻印された試薬チューブ、およびこのサンプル・チューブが96本まで格納出来る試薬ラックによって管理される。

サンプル・チューブと同様に、試薬ラック内の各試薬チューブの配置はバーコードをスキャナーで読み取り、この配置をデータとして管理され、試薬ラック自身にもバーコードが添付される。

プレート：プレートには、マスター・プレート、試薬プレートおよびアッセイ・プレートの3種類がある。

マスター・プレートは1枚につき96のウェル（サンプルや試薬を分注するプレート上の窓）を持ち、同じく最大96のサンプル・チューブが収納可能なサンプル・ラックからサンプルが分注される。この時、1つのサンプルは1個所のウェル…ラック内のサンプル・チューブのレイアウトと同じポジションのマスター

・プレートのウェルに分注される。この為、マスター・プレートはサンプル・ラックの「子」のデータとして管理される。

試薬プレートも（マスター・プレートと同様に）96のウェルを持つが、1枚の試薬プレートには1種類の試薬のみを使うので、試薬プレートは試薬チューブの「子」のデータとして管理される。試薬チューブの試薬ラック内の配置情報を保持する理由は、分注に使う自動機器が、ラックに格納した試薬チューブの使用を要求するからである。

アッセイ・プレートは384のウェルを持ち、最大96のサンプルに対して4SNPのタイピングを同時に行う事が出来る。この1枚のアッセイ・プレートは、データ上4つの仮想的なバーチャル・プレートとして管理される。1つのバーチャル・プレートは最大96のサンプルに対して1つのSNPのタイピングを行うプレートで、これが仮想的に4枚合わせり1つのアッセイ・プレートを構成する。

図3に、基本的なサンプル・試薬チューブ、各種プレートおよび関連したデータの基本的なデータ・スキマを示す。

## 2) 「ハーディ、ワインバーグ平衡」を用いたタイピング・データの検定

SNPでは、その座位の塩基が2つのバリエーション、例えばA（アデニン）またはG（グアニン）の様な2つのアレル（対立遺伝子）を持つ。全ての染色体は対になっているので、SNP座位も対になっている染色体の各々に一つずつの合計2箇所に存在する。この為、TaqManアッセイ等で観測されるSNPのパターンは、1) 一方アレルにの場合、A) のホモ接合体であるA-Aと、2) これとは異なるアレル（G）のホモ接合体G-G、3) 対立するアレルが一つずつ組合わさったヘテロ接合体A-G、以上の3つパターンになる。

一方の遺伝子上でそのSNPがAを持つ確率を $\alpha$ とすれば、A-Aホモ接合体の確率は $\alpha^2$ であり、G-Gホモ接合体の確率は $(1-\alpha)^2$ 、A-Gヘテロ接合体の確率は $2\alpha(1-\alpha)$ となり、この関係を「ハーディ、ワインバーグ平衡」と呼ぶ。

「ハーディ、ワインバーグ平衡」が成立する条件は、多くの世代の交配を経て「平衡」した状態にあるサンプル集団から無作為に、同時に統計的に「平均」に抽

出されなければならない。

ここで、そのサンプルのデータが「ハーディ、ワインバーグ平衡」から大きく離れた値を取った場合には、1) そのデータを生成したアッセイ自体がコンタミネーション等によって「汚染」されたデータである、2) タイピングされたサンプル集団が統計的には「平均」に抽出されていない、の何れかに相当すると考えられる。

S N P 機能解析の目的は、或る形質の発現（疾病等）の有無によって区分されるサンプル・グループの間で S N P データを比較して、その形質を発現させる要因となる S N P を特定する事である。つまり、要因となる S N P をタイピングした場合、そのグループを特徴付ける形質と因果関係のある S N P を「平均」以上に持っている為、「ハーディ、ワインバーグ平衡」から離れたアレルの分布を持つはずである。言葉を代えれば、この2) に該当する S N P を見つけることこそ、S N P 機能解析および本プロセス・モデルの目的である。

「目標」 S N P を特定する為に、「ハーディ、ワインバーグ平衡」から逸脱したアレル分布の S N P から、1) の「汚染」されたデータを持つ S N P を除外する。

このコンタミネーション等の「汚染」されたデータを識別する方法が、1枚のアッセイ・プレートで4つの S N P をタイピングする方法である。図4に、アッセイ・プレート上の S N P の配置パターンを示す。

コンタミネーションや P C R の不調等のアッセイ自体の失敗が起きた場合、同じプレート上の他の S N P についても異常が発生すると考えられる。確率的には、要因となる S N P を偶然複数選び出す事は極めて稀と考えられるので、同一プレート上で複数の S N P データが「ハーディ、ワインバーグ平衡」から逸脱したケースは、タイピングの失敗と判断しそのプレートで行った全データの破棄（とアッセイのやり直し）を行う。

更に、タイピングした S N P 同士が連鎖不平衡する可能性を回避する為、4つの S N P を可能な限り離す=「走査領域」を4つの区間に分け、その各々の区間からタイピングする S N P を1つ選択する。

#### 段階4：タイピング・データによるハプロタイプの解析

この工程では、ハプロタイプと連鎖不平衡の2つのコンセプトを用いて、「目標」SNPの近傍のSNP（最終的に「目標」SNP自身）を特定する。

一つの配偶子（対となる染色体の一方）上にある対立遺伝子（SNPのアレル）の組合せであるハプロタイプは、TaqMan法等のSNPタイピング・アッセイで得られたデータを基に、確率的に予測される。

このハプロタイプを確率的に予測する事について、AまたはGを取るSNP#1、TまたはGを取るSNP#2、TまたはCを取るSNP#3、以上の3つのSNPが取るハプロタイプのケースから考える。

例えば、サンプルXで、SNP#1がA-Aのホモ接合体、SNP#2がT-Gのヘテロ接合体、SNP#3がT-Cのヘテロ接合体を取る場合には、以下の2つのケースで4通りのハプロタイプの存在が予測される。（この例の各ハプロタイプの確率は各々25%となる。）

ケース1)	<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>確率</u>
染色体1	A	T	T	25%
染色体2	A	G	C	25%
ケース2)	<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	
染色体1	A	T	C	25%
染色体2	A	G	T	25%

更にもう一つのサンプルYが、SNP#1がA-Aのホモ接合体、SNP#2がT-Gのヘテロ接合体、SNP#3がT-Tのホモ接合体を取る場合には、各々50%の確率を持つ2通りのハプロタイプが予測される。

ケース1)	<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>確率</u>
染色体1	A	T	T	50%
染色体2	A	G	T	50%

そして、この2つのサンプルから予測されるハプロタイプを以下の通りとなる。

<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>確率</u>
A	T	T	25%+2+50+2=37.5%

A	T	C	$25\%+2=12.5\%$
A	G	T	$25\%+2+50+2=37.5\%$
A	G	C	$25\%+2=12.5\%$

実際の解析は、数個から数十程度の範囲で定められた一定の S N P 数含む連續した領域をウインドと定義し、そのウインド内の S N P のタイピング・データ（全サンプル）から、ハプロタイプの組合せとその各々の出現する確率を統計的に求める。このウインドに含まれる S N P の数が多すぎるとハプロタイプ個々の確率が低くなり、連鎖もしくは連鎖不balanceの有無の確認が微妙になるので、10程度の S N P を含むウインドの定義が有効である。

図2は、このウインドをそのサイクルにおける「走査領域」の先頭から終端まで移動させて、そのウインド内に含まれている S N P データを解析するイメージを表している。

#### 段階5：「マーカー」S N P の推定

この工程では、図5に示すように、特定のハプロタイプの確率が突出している領域を見つけ出す事によって、「目標」 S N P の近傍領域の推定を行う。

段階4で解析されたハプロタイプのデータから、解析された S N P グループ間で連鎖不balanceが見られたか否かの判定が可能となる。

これらの S N P の間に連鎖不balanceが見られない場合、各 S N P のアレルの出現頻度は「平均的」な値に落ち着き、加えて、各々の S N P が「独立」している為、そこから統計的に求められるハプロタイプについても、特定のハプロタイプに集約されない、広く薄く分散したものになると考えられる。

これに対し、解析された S N P グループ間で連鎖不balanceが見られたる場合、そこにはサンプル・グループを統計的に特徴付ける S N P が含まれている事であり、それらの S N P では特定のアレルの出現頻度が増大する。そして、これら S N P データを統計的に解析した結果であるハプロタイプの確率分布も、（「目標」 S N P をアッセイしなかった場合より）特定のハプロタイプに集中する事が予測される。

この特定ハプロタイプへの集中を識別する方法として、個々のハプロタイプの出

現頻度を比較するほかに、その解析データから予測されるハプロタイプの総数、これらハプロタイプの標準偏差、確率上位のハプロタイプ・グループのハプロタイプ全体に対する出現頻度の割合を「統計量」として観測し、作用・副作用の有無という形質の発現によって区分されるサンプル・グループ間でこれらを比較する。

しかしながら、先にも述べた様に、「目標」SNPを選び出して直接タイピングする事はなかなか困難である。この問題の解が、「目標」SNPが近傍のSNPと連鎖不平衡する事を応用し、「目標」SNPの近傍領域を推定する事である。連鎖不平衡した近傍のSNPは、「目標」SNPを直接解析した場合と比べれば程度こそ弱いが、やはり同じようにハプロタイプの確率分布が変化する事が期待される。このような近傍のSNPは、「目標」SNPの「マーカー」SNPであると考えられる。

直接「目標」SNPをタイピングする事無しに、「マーカー」SNPのタイピング・データから、特定ハプロタイプの突出を示す領域を見つけ出し、その領域を次の「走査領域」として、段階2から再度工程を繰り返す。（図6を参照）又、そのサイクルで、「走査領域」中の全てのSNPがタイピングされているのであれば、次の段階6へ進んで、その特定ハプロタイプに収束しているSNP座位が「目標」SNPとして決定される。

この段階の目的は「マーカー」SNPの推定であり、「走査領域」を新しく推定された「マーカー」SNPの近傍に絞り込むサイクルを繰り返す事で、「マーカー」SNPを「目標」SNPを接近させる事と言える。（最終のサイクルの以前に「目標」SNPが解析から導き出されていたとしても、その時点ではまだそのSNPを「目標」SNPとして確定出来てはいないので、「マーカー」SNP。

)

#### 段階6：「目標」SNPの特定

この工程の目的は、全工程で選び出された「目標」SNPと開発新薬の薬剤応答性との相関であり、その度合いを定量的に導き出す事である。

この工程では、「目標」SNPとして確定されたSNPのアレル頻度について、サンプル・グループでの最終的な検定を行う。（カイ2乗検定や最尤法等による

)

更に、サンプル・グループ以外のS N P データ、特に東大医科研が所有する日本人の標準的なS N P に関するデータとの比較を行う事も効果的である。又、一部のゲノム研究企業で、S N P に関連するデータベースを販売する動きが有るので、これらの利用も可である。

カイ<sup>2</sup>乗検定や相関解析を用いて、「～パーセントの確率で期待される効果が得られる」、または「～パーセントの確率で重篤な副作用が発生する」等の評価検定を行う。

#### 【0006】

##### 【実施形態の効果】

この実施形態によれば、タイピングされたS N P の中から疾患易罹患性または薬剤応答性に関連するS N P を特定する際に、解析の対象となる塩基配列領域を大きな領域からより局所的な領域へと段階的に絞り込みを行う事で、加えて、品質管理された工程でS N P のタイピングを行う事によって、最終的にこれら関連するS N P を特定することが出来る。

#### 【0007】

##### 【他の実施形態】

段階1の「走査領域」の決定では、初期の「走査領域」を、大まかな染色体レベルより詳細な遺伝子レベルで設定することも可能である。これによって、狭い初期「走査領域」からタイピングが開始できるので、より効率的な「走査領域」が可能となる。

段階2のタイピングS N P の決定では、(特に、早い段階の「走査領域」において)観測するS N P 座位の間隔が大きすぎた場合、連鎖不平衡はほとんど観測する事が出来ない。この連鎖不平衡が観測可能なレベルの物理的距離より短い間隔でタイピングS N P を選ぶ事が必要である。

初期の「走査領域」では、この間隔を、1000 S N P について1つのS N P を選ぶ=10万塩基について1つのS N P を選ぶように初期値が設定される。しかし、これはあくまで計算値であり、その近傍に見られる連鎖不平衡の度合いによっては、より小さい値(数十S N P について1つのS N P を選ぶ)を取るケース

も有る。

段階5の「マーカー」SNPの推定において、ハプロタイプの総数や出現頻度等の「統計量」の変化を比較する以外の方法として、確率上位のハプロタイプが共通SNPパターンを求めることで連鎖もしくは連鎖不平衡の有無を確定することができる。

先の【発明の実施の形態】の項で上げたハプロタイプの確率の例では、サンプルXとサンプルYのハプロタイプの確率の和は以下の通りであった。

<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>確率</u>
A	T	T	$25\%+2+50+2=37.5\%$
A	G	T	$25\%+2+50+2=37.5\%$
A	T	C	$25\%+2=12.5\%$
A	G	C	$25\%+2=12.5\%$

これを、確率上位の順から並べ直すと次のようになる。

<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>確率</u>
A	T	T	50%
A	T	C	25%
A	G	T	12.5%
A	G	C	12.5%

この例では、このハプロタイプのグループは、75%の確率でSNP # 1がAのアレルSNP # 2がTのアレルを取る事を示し、SNP # 1とSNP # 2に連鎖が有ったものと推測される。

より複雑な下記の例では、

<u>SNP#1</u>	<u>SNP#2</u>	<u>SNP#3</u>	<u>SNP#4</u>	<u>SNP#5</u>	<u>確率</u>
A	T	T	A	T	30%
A	T	T	A	G	25%
C	T	T	A	T	10%
C	T	T	A	G	10%
A	T	T	C	T	5%
A	T	T	C	G	5%
A	G	C	A	T	3%

55%の確率で、SNP # 1がアレルA、SNP # 2がアレルT、SNP # 3がアレルT、SNP # 4がアレルAを持つハプロタイプのパターンが観測される。

更に、75%の確率でSNP # 2がアレルT、SNP # 3がアレルT、SNP # 4がアレルAを持つハプロタイプのパターンが、85%の確率でSNP # 2がアレルT、SNP # 3がアレルTを持つハプロタイプのパターンが観測される。これによって、このサンプル・グループのデータからはSNP # 2とSNP # 3を中心とした連鎖が見られるものと判断される。

ここで、連鎖の有無を分ける出現確率の閾値は、ウィンドで走査するSNPの数に関連する。SNP数が増えるとハプロタイプのバリエーションが増加して個々のハプロタイプが観測される確率が減るので、その結果、閾値の値も低くなる。

10のSNPを観測するウィンドを使った場合、閾値としては70%が適当と考えられる。

この方法を用いる事で、比較する2つのサンプル・グループ間に存在する異なる連鎖、もしくは一方のサンプル・グループにしか存在しない連鎖を確認する事で、「目標」SNPまたはその近傍の領域を特定する。

#### 【0008】

##### 【発明の効果】

以上説明したように、この発明によれば、マーカーとなるSNPを推定する事で解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞り込みを行い、更に、品質管理された工程でSNPのタイピングを行う事によって、最終的に、疾患易罹患性や薬剤応答性等に関連するSNPを特定するこ

とが出来る

【図面の簡単な説明】

【図1】

従来のS N P機能解析のプロセス・フロー、および疾患易罹患性または薬剤応答性に関連するS N P特定のフロー、以上の2つを図示。

【図2】

本プロセス・モデルにおける、「走査領域」の決定から、ハプロタイプ解析の実行までの手順を図示。

【図3】

チューブ、プレートおよびこれらに関連する基本的なデータのデータ・モデルを図示。

【図4】

384-ウェルのアッセイ・プレートでタイピングされるS N Pのレイアウトを図示。

【図5】

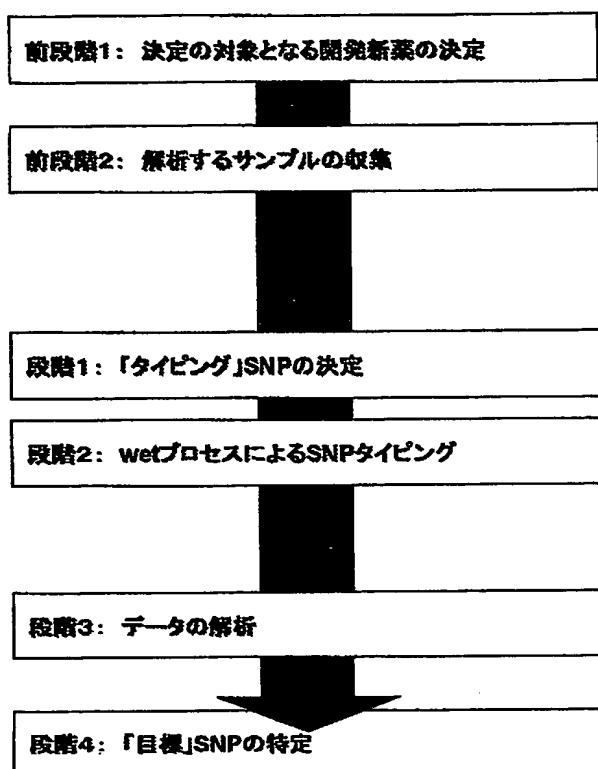
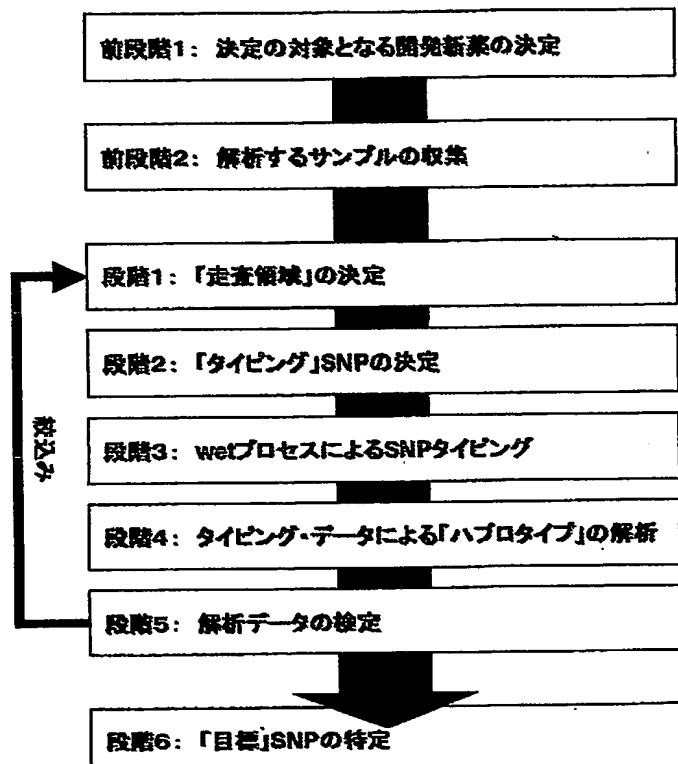
段階5の解析データの検定の前半部分にあたる、ハプロタイプ解析から統計量に変化が現れた領域を識別する手順を図示

【図6】

段階5の解析データの検定の後半部分にあたる、新しい「走査領域」を定める手順を図示。

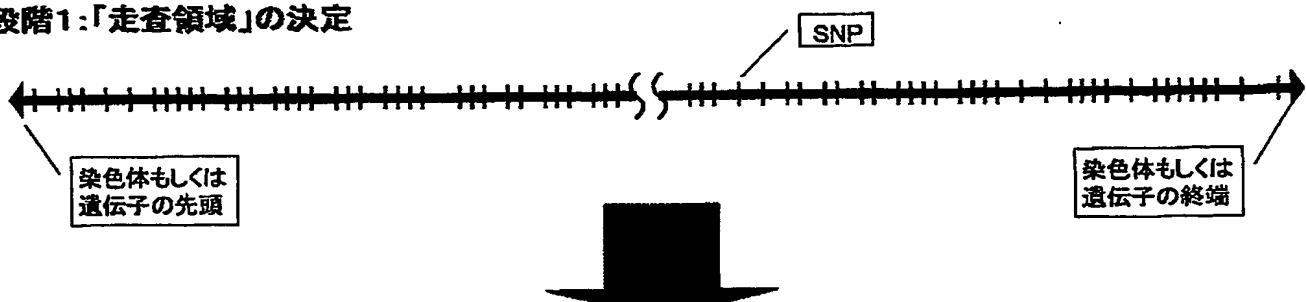
【書類名】 図面

【図1】

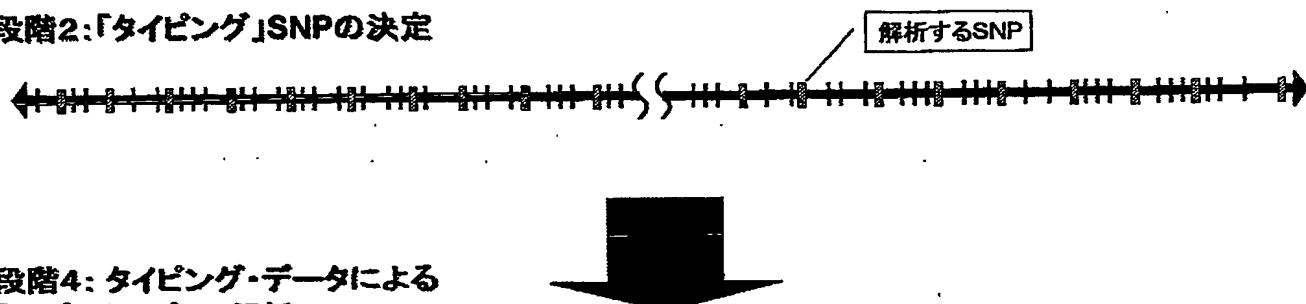
従来のSNP機能解析のプロセス・フロー疾患易罹患性または薬剤応答性に関するSNP特定のフロー

【図2】

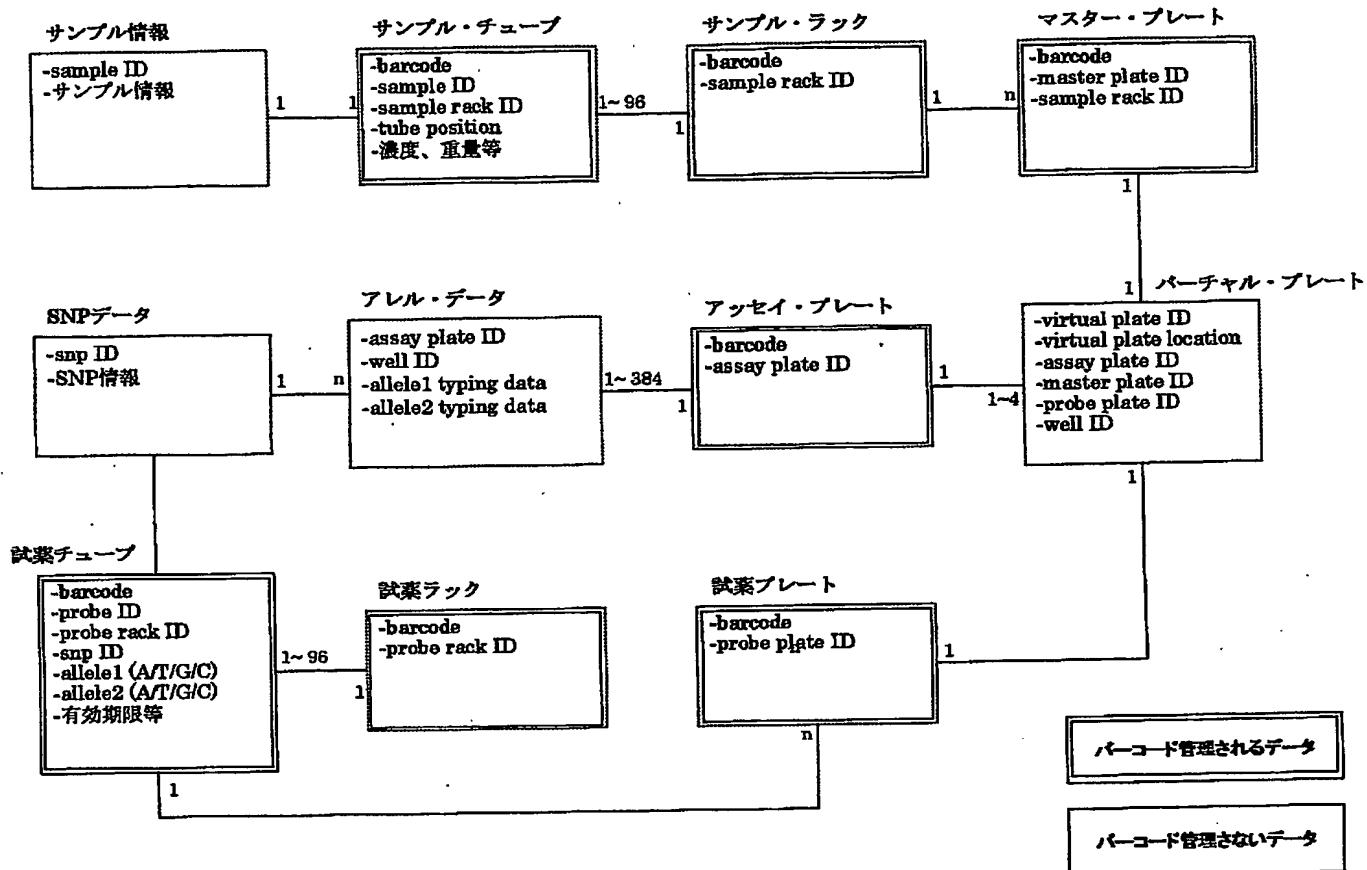
## 段階1:「走査領域」の決定



## 段階2:「タイピング」SNPの決定

段階4: タイピングデータによる  
「ハプロタイプ」の解析

【図3】



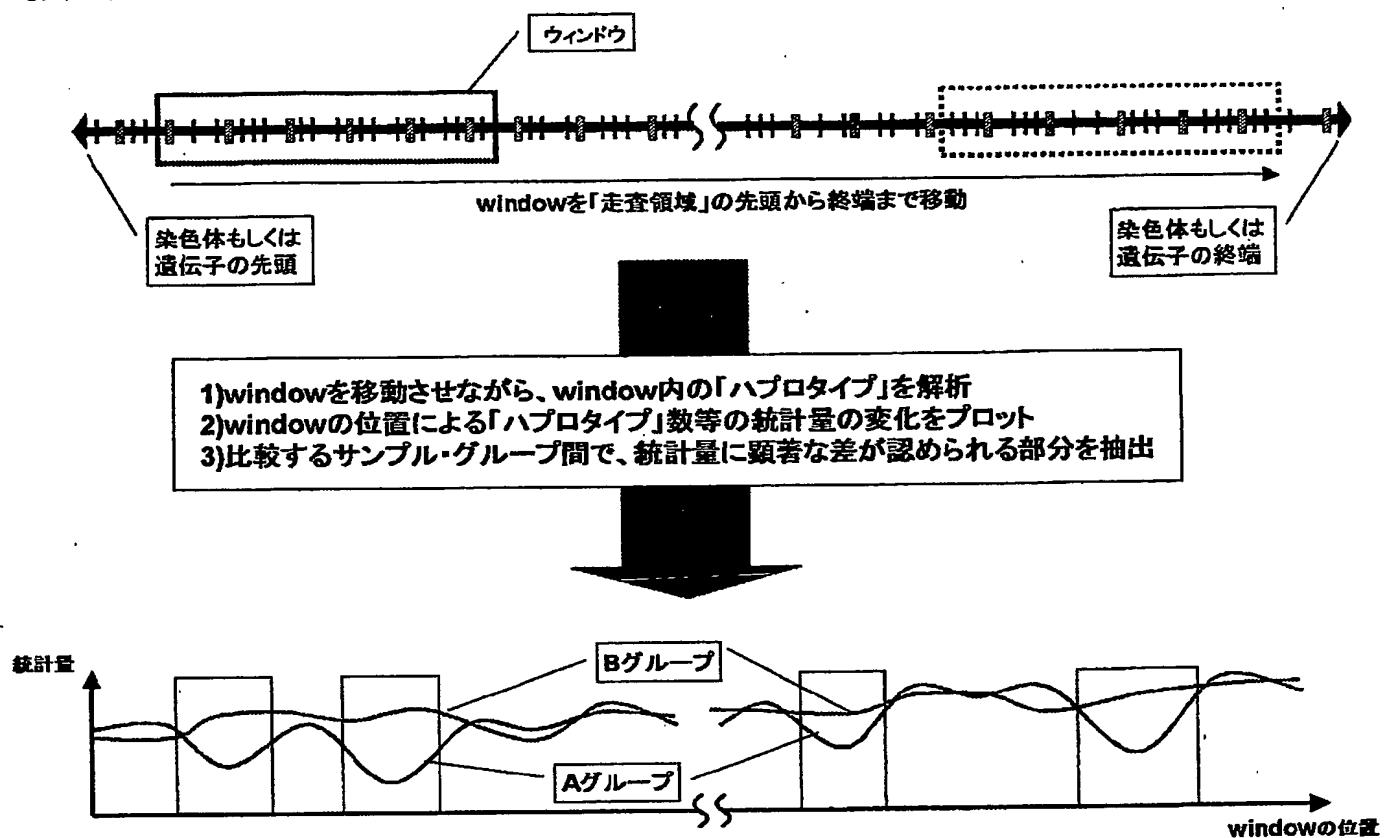
【図4】

384ウェルのアッセイ・プレート上のSNPの配置

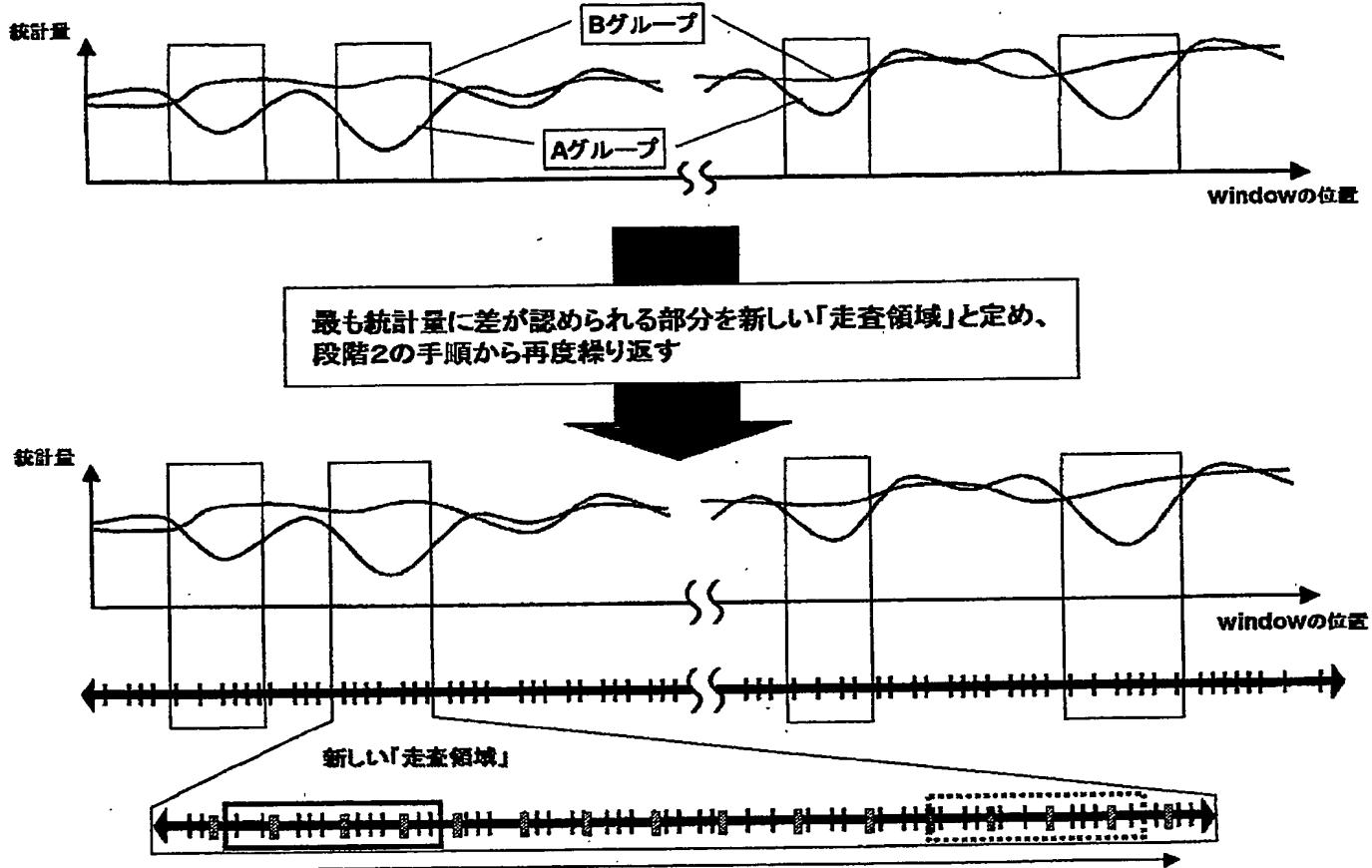
右図の(1)(2)の4つのウェルのブロックで、1つ  
のサンプルに対し4種類のSNPをタピング

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	(1) (2)															
2	(3) (4)															
3	(1) (2)															
4	(3) (4)															
5	(1) (2)	(1) (2)														
6	(3) (4)	(3) (4)														
7	(1) (2)															
8	(3) (4)															
9	(1) (2)															
10	(3) (4)															
11	(1) (2)															
12	(3) (4)															
13	(1) (2)															
14	(3) (4)															
15	(1) (2)															
16	(3) (4)															
17	(1) (2)															
18	(3) (4)															
19	(1) (2)															
20	(3) (4)															
21	(1) (2)															
22	(3) (4)															
23	(1) (2)															
24	(3) (4)															

【図5】



【図6】



【書類名】 要約書

【要約】

【課題】 既存のS N P機能解析プロセスの応用では、文献や既存ゲノム関連データベース等の限られた情報から、タイピング・解析の為のS N Pグループを効果的に選び出す事は難しく、更に、コンタミネーション等のデータ品質低下の原因を排除する事も困難である。

【解決手段】 S N Pデータを用いた疾患易罹患性または薬剤応答性を特定方法において、「走査領域」を、大まかな領域からより高い局所的な領域へと段階的に絞り込みながら、最終的に「目標」S N Pを特定する。更に、バーコードによるサンプル・プレートの管理、タイピング・データの検定を行う事によって、データ品質を保証する。

【選択図】 図1

出願人履歴情報

識別番号 [502060360]

1. 変更年月日 2002年 1月15日

[変更理由] 新規登録

住 所 神奈川県藤沢市川名849の10クリオ藤沢参番館506  
氏 名 鈴木 俊明

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER: \_\_\_\_\_**

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.